# Phonetic characteristics of filled pauses: the effects of speakers' age

Mária Gósy[1], Judit Bóna[2], András Beke[1], Viktória Horváth[1],

[1]*Department of Phonetics, Research Institute for Linguistics, HAS, Budapest, Hungary*
[2]*Department of Phonetics, Eötvös Loránd University, Budapest, Hungary*

gosy.maria@nytud.mta.hu, bona.judit@btk.elte.hu, beke.andras@nytud.mta.hu, horvath.viktoria@nytud.mta.hu

## Abstract

Filled pauses usually reveal speech planning or execution problems even though the speaker does not produce an overt error and may have a function of discourse marker as well. In Hungarian, the most frequent form of filled pauses is a schwa-like vowel of various durations. The purpose of this study was to analyze the occurrence, duration and formant structure of Hungarian schwa-like filled pauses in 16 nine-year-old children, in 16 young adults and in 16 elderly speakers. Results confirmed age-dependent patterns of filled pauses. Speakers' age is one of those factors that influence the occurrences and formant values of filled pauses.

**Keywords**: duration of filled pauses, formant structure, children, young adults, elderly

## 1. Introduction

Filled pauses are known also as mazes, delay markers, hesitation (phenomena), editing pauses, unlexicalized filled pauses, fillers, interjection, delays, even noises, and they usually reveal speech planning or execution problems even though the speaker does not produce an overt error (Shriberg 2001, Watanabe et al. 2008). The speaker might use filled pauses to gain time for synchronization between thinking and speaking processes. In addition, filled pauses also reflect the speaker's strategy to control his/her own speech production and they can have pragmatic and discourse functions as well (Clark and Fox Tree 2002, Simpson 2006). The term 'filled pause' will be used in this paper as the phenomenon containing a schwa-like vowel in Hungarian with the same functions like *uh*, *uhm* or *er* in English.

The main theoretical interest that guided our research was to find out whether the durations and the formants of Hungarian filled pauses are the same or specifically different across ages. The effects of age on filled pauses are not completely understood and, in fact, they are vigorously debated in the literature (e.g., Gayraud et al. 2011). Filled pauses in large corpora were reported to be from one-third to over one-half of all disfluencies (Shriberg 1994). Most studies agree that there are no large differences in the frequency of disfluencies between young and adult speakers (e.g. Leeper and Culatta, 1995), even centenarians' breakdowns (frequency and types of disfluencies) were similar to those of younger elderly speakers (Searl et al. 2002). On the contrary, elderly people were reported to use a larger number of filled pauses as opposed to young adults (e. g. Kemper 1992, Roggia 2012). Filled pauses are reported to occur in children's spontaneous speech as early as in three-year-olds (e.g. Furman and Özyürek 2007, Hudson Kam and Edwards 2008). Analysis of children's spontaneous speech authors concluded that it shows more adult-like disfluency patterns toward the later preschool years (DeJoy and Gregory 1985). In addition, there is a continuous decrease in the occurrences of disfluencies from 6-year-olds to adulthood, however, the proportions of diverse types of disfluencies show differences depending on age (Ito 1986). Nine-year-old Turkish-speaking children were shown to use filled pause (in form of *şey*) as a planning/hesitation marker and a narrative initiation marker in 62% of all items while they occurred as fillers in 35% of the cases (Furman and Özyürek 2007). The close interaction between the occurrence of filled pauses (and other breakdowns) and grammatical structures of speech samples in children were confirmed in several studies (e.g. Fiestas et al. 2005, Farantouri et al. 2008). In children, the frequency of filled pauses increases with linguistic complexity on the one hand, being greater in longer utterances (Yaruss et al. 1999, Thordardottir and Weismer 2002).

Filled pauses are reported to have wide durational range from less than 100 ms to about 750 ms (e.g. Shriberg 2001, Duez 2001, Eklund 2004) but they may frequently have even longer durations. The duration of filled pauses shows a great variety also in children. The average duration was found to be about 800 ms in 4.6-year-olds' spontaneous speech samples (MacWhinney–Osser 1977) while nine-year-old Spanish-speaking children had 28% filled pauses of their all analyzed disfluencies (Esposito 2005). There was no significant difference among 3-year-old, 5-year-old and 9-year-old Turkish-speaking children's speech samples in the occurrence of filled pause *şey* (Furman and Özyürek 2007). Spanish-speaking six-year-olds' narratives contained filled pauses in 2.3% (on average) of all produced words while English-speaking children had them in 3.9% of all produced words (Fiestas et al. 2005). Our hypotheses were that (i) filled pauses would show age-dependent occurrences and (ii) age would influence both the durations and formant structures of filled pauses.

## 2. Methodology

Spontaneous speech samples were analyzed in three groups of: (i) 16 nine-year-old children, (ii) 16 young adults (ages 22 to 28) and (iii) 16 elderly people (ages 75 to 90). All speakers were native speakers of Hungarian from Budapest. Half of them were females while the other halves were males in all age groups. Children were randomly selected from three elementary schools. No hearing or speaking disability was reported among them. Adults were randomly selected with the criteria of their ages and gender from BEA, a large Hungarian spontaneous speech database (Gósy 2012). Their hearing and speech production was appropriate for the ages.

The topics of the adults' narratives included the speakers' families, work or past work, and hobbies while children spoke about family, school, holiday and hobbies. An average of 8.5 minutes was selected from the middle of each narrative. The samples collected from children were carried out at school in a small quiet room in the morning.

Speech samples of all subjects were manually annotated by two of the authors using Praat (Boersma and Weenink 2010). Each filled pause was identified and coded by two of the

authors while the other two authors controlled the coding. No disagreement was found in identification of the filled pauses in speech samples. Functions of filled pauses were neither identified, nor classified. 71.5% of all filled pauses were schwa-like (neutral) vowels.

Altogether 1054 schwa-like filled pauses were analyzed in the three groups (249 with children, 523 with young and 282 with elderly participants). The occurrences, durations, and the first two formants of the filled pauses were analyzed. Their duration was measured from the first glottal pulse to the last glottal pause of the vowel. The first two formants were measured at the midpoint of the total filled pause duration. Both duration and formant frequency measurements were carried out using an automatic Praat script followed by their manual verification using auditory feedback if necessary. The script was written for the purpose of this study by one of the authors.

MANOVA and non-parametric tests (Kruskal–Wallis and Mann–Whitney test) (SPSS, version 19.0) as well as three-dimensional Euclidian distance measures were used to test the hypotheses.

## 3. Results

This study concentrates on measured occurrences, durations and formants of schwa-like filled pauses in Hungarian, where they typically appear as *Ö* ([ø] or [ə]) with various durations. 71.63% of all filled pauses in the present material were produced by this vowel while the rest of them consisted of a combination of a similar vowel and a [h]-like and/or a nasal-like consonant or a murmur, for example *mm*, *Öh*, *Öm*, *Öhm*. For example (filled pause is marked by bold **Ö** letters): *nekem mondjuk* **Ö** *én mondjuk fel akarok használni* 'for me let say **Ö** let say I want to use'. The ratio of schwa-like filled pauses was 78.5% in young adults, 67.6% in elderly and 68.8% in children.

### 3.1. Occurrences of schwa-like filled pauses

The occurrence of schwa-like filled pauses showed no difference between children (2.7 indices/minute, on average) and elderly participants (2.6 indices/minute, on average) but it was more frequent in young speakers (2.9 indices/minute, on average). Females used filled pauses more frequently (435 indices) than males (487 indices); however, there is almost no difference between nine-year-old girls (122 indices) and boys (125 indices). Both young and old female adults used more filled pauses (2.99 indices/minute and 3.5 indices/minute) than young and old male adults (2.49 indices/minute and 1.6 indices/minute).

### 3.2. Durations of schwa-like filled pauses

Young adults produced the longest filled pauses while children produced the shortest ones while the duration of elderly's filled pauses fell within those of the two former age groups (Table 1).

Table 1: *Occurrence of filled pauses depending on age (M=mean, SD= standard deviation).*

| Age groups | Duration of filled pauses (ms) | | | | | |
|---|---|---|---|---|---|---|
| | all speakers | | females | | males | |
| | M | SD | M | SD | M | SD |
| 9-year-olds | 269 | 110 | 276 | 100 | 270 | 123 |
| young adults | 340 | 163 | 275 | 137 | 350 | 163 |
| elderly | 322 | 163 | 280 | 139 | 322 | 163 |

The differences depending on age is significant (Kruskal–Wallis test: Chi-Square (2) = 10.813, $p = 0.004$). There was no significant difference between nine-year-old girls and boys (according to Mann–Whitney test, $p > 0.05$). Young adults' and elderly's data, however, were proved to be significantly different depending on gender (Mann–Whitney test for young adults: $Z = -4.947$, $p < 0.001$ and for elderly: $Z = -2{,}109$; $p = 0.035$), see figure 1.

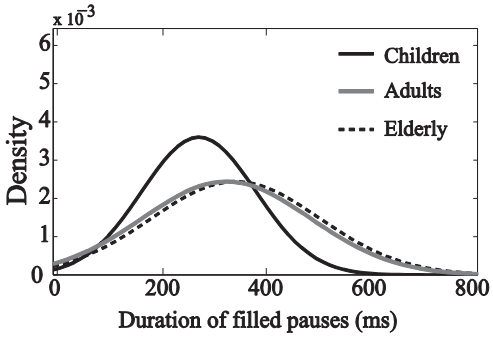


Figure 1: *The interrelations of duration and relative frequency of filled pauses depending on age.*

Nine-year-old boys' filled pauses were significantly shorter than those of young and old male adults (Kruskal–Wallis test: Chi-Square (2) = 24.5453, $p < 0.001$), and both adults groups' durations differed significantly from those of the nine-year-old boys (Mann–Whitney test between boys and young adults: $Z = -4.982$; $p < 0.001$; and $Z = -2.941$; $p = 0.003$ between boys and elderly males). No statistical difference was found between the data of the two male adult groups. Although elderly female speakers produced some longer filled pauses than young females and nine-year-old girls, the differences did not turned out to be significant.

### 3.3. Formants of schwa-like filled pauses

Formant frequency data were analyzed considering the factors of 'age' and 'gender' (Figure 2). Both first and second formants show differences depending on age. F1-values decrease in parallel with age with both females and males. The F2-values decrease between children and young adults both in females and males but increase between young and elderly speakers. The decrease of the F1s is assumed to reflect the slight rise of the tongue height across ages. The increase of the F2s in elderly's filled pauses might be assumed to show the consequences of either the higher tongue position in the oral cavity or the less lip-rounding (or both). This modification seems to be characteristic particularly of elderly males.

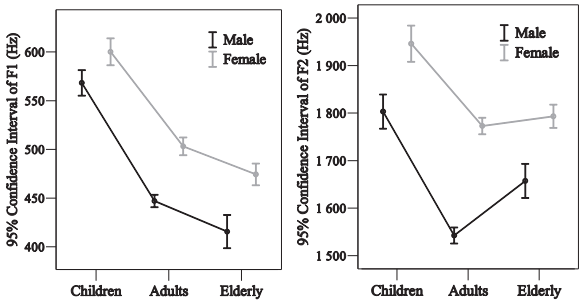


Figure 2: *Formants of filled pauses depending on age and gender.*

The effect of 'age' and 'gender' was statistically analyzed using MANOVA and showed significant effects on the formants (factor of 'age' for F1: $F(2, 1044)= 277,985$, $p < 0,001$, $\eta^2$ partial = 0.35; for F2: $F(2,1044) = 141,686$, $p < 0,001$, $\eta^2$ partial = 0.21 and factor of 'gender' (for the first formant: $F(1, 245) = 11.009$ at $p < 0.001$, $\eta^2$ partial = 0.094; for the second formant: $F(1, 245) = 29.269$, $p < 0.001$, $\eta^2$ partial = 0.018). The interaction of 'age' and 'gender' was statistically significant but only for the second formant ($F(2, 1044) = 9,571$, $p < 0.001$). The first two formants were modeled by Gauss distribution in order to demonstrate the differences among the three age groups. Children's formants differed largely from both adult groups' while the two adult groups' formant values showed larger overlaps. Tukey post hoc tests revealed that the formant frequency values of the two formants differed from each other in all age groups ($p = 0.0001$ in all cases).

As expected, the factor 'gender' showed significant differences in both adult groups for both F1s and F2s (one-tail ANOVA in the case of young adults for first formants: $F(1, 519) = 99,606$, $p < 0.001$; and for second formants: $F(1,519) = 347,479$, $p < 0.001$) and in the case of elderly for the first formants: $F(1, 274) = 35,126$, $p < 0.001$; and also for the second formants: $F(1,274) = 40,331$, $p < 0.001$). Although the girls' first and second formants showed less differences from those of the boys' than young adults had, the data turned out to be significantly different even in their cases as well (for the first formants: $F(1, 245) = 11,009$, $p < 0.001$; and for second formants: $F(1,245) = 29,269$, $p < 0.001$).

The first two formants were modeled by Gauss curve in order to demonstrate the differences among the three age groups. The distance between the groups was characterized in the model by means of the Euclidian distance that was calculated based on the mean value of the distributions of the first two formants. Results confirmed that children's formants differ largely from both adults groups' while the two adult groups' formant values show larger overlaps.

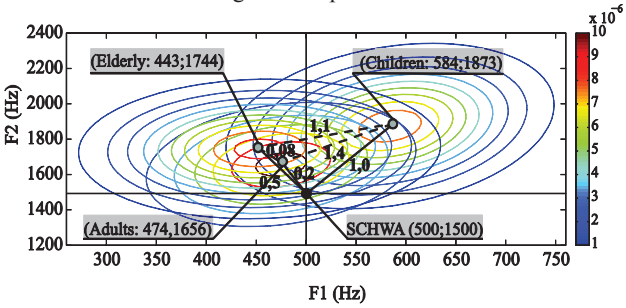


Figure 3: *The Euclidian distance of the formants of filled pauses from those of the (classical) neutral vowels.*

We wanted to define the closest filled pause formants to those of the neutral vowel among the age groups. The F1 value of the neutral vowel is 500 Hz while that of the F2 is 1500 Hz (Pickett 1980). The distance of our data from those of the neutral vowel was calculated by means of Euclidian distance from the mean values of the distributions. Results indicate that the formants of the young speakers' filled pauses are the most similar to those of the neutral vowel, elderly's formants are less similar while those of children's show large differences (Figure 3).

## 4. Discussion and conclusion

Our study aimed to investigate the occurrence, durations and formants of Ö-type filled pauses in Hungarian. The finding that filled pauses occurred most frequently in young adults as opposed to children and elderly. There was practically no difference in occurrences of filled pauses between our 9-year-old girls and boys. Hungarian-speaking elderly subjects' narratives contained the least amount of filled pauses which contradicts some findings reported in the literature (e.g. Roggia 2012).

Elderly speakers are frequently reported being as fluent as, or sometimes more fluent than, young speakers, and filled pauses, were shown to increase with age (e.g. Leeper and Culatta 1995, Pindzola 1990). The frequency of filled pauses in our material seems to contradict those claims that occurrence of filled pauses increases in elderly (see Roggia 2012).

The hypothesis that filled pauses would show similar durations across ages was partly confirmed. Analysis of the durations disclosed that the difference across ages exists between nine-year-old boys and adult males where children had the shortest filled pauses and young males the longest ones. No such difference exists among females, the durations of the filled pauses are very similar irrespective of age. The durational difference of the filled pauses seems to be rather a factor of gender than a factor of age.

Our study confirmed that 'age' and 'gender' are decisive factors for the first two formants of the Hungarian schwa-like filled pauses. The elderly participants' formant values overlap with those of young adults' as a succeeding that the formants of the schwa-like filled pauses show the acoustic consequences of the articulation in the mid part of the oral cavity. Our findings seem to partly contradict the respective facts reported in the literature. Children's vowel formants are reported to be higher than those of adults' (Busby et al. 1995) because of the different shapes and sizes of their speech organs including the developing vocal cords (Baken–Orlikoff 2000). Elderly's physiological configurations, somewhat centralized articulation, and motor control of their articulation may explain the acoustic differences in their vowel formants (Hooper–Cralidis 2009).

The effect of gender on formants yielded significant differences in all our age groups. There are contradictory findings in the related literature concerning the gender differences appearing in the formant values in children. No differences depending on gender were found in children's speech until the age of 15 in a study by Lee et al. (1999) while the opposite was confirmed in the study by Busby and Plant (1995) claiming that formant values were higher in females than in males (children's ages varied between 5 and 11). Vorperian and Kent (2007) found gender differences in formant frequency values by age 4 with more apparent differences by age 8. Since children's speech organs, including vocal cords, can be considered to be identical at the age of 9 (Simpson 2009) therefore no large gender differences were assumed for the formants of the filled pauses in our children. Although supraglottal anatomy and the vocal tract configurations (particularly vocal tract length and mouth opening size) are relevant for formant frequencies; however, other factors of learned behaviors and the vowel type should also be considered in gender differences of speech beside the age (Busby and Plant 1995, Simpson 2009). We can conclude that by the age of 9 children are expected to show gender differences in schwa-type Hungarian filled pauses. As expected, female formants were in general higher than those of male formants, and the vowel space defined by F1 and F2 is shown to be larger in females than in males irrespective of age (see also Simpson 2009).

Our findings partly supported the claims about the occurrence and durations of filled pauses reported in the literature. The

differences found can be explained primarily by the specific articulation of the Hungarian filled pauses. The vowel-like sound constituting filled pause in Hungarian is the realization of two vowels (the front, labial, mid vowel and the neutral vowel) that are both members of the Hungarian vowel inventory but they have different phonemic status.

## 5. Acknowledgements

This study was supported by OTKA project No. 108761, and the Bolyai János Research Scholarship.

## 6. References

Baken, Ronald and Robert F. Orlikoff (2000). *Clinical measurement of speech and voice*. San Diego: Singular Publishing Group, 594 p.

Boersma, Paul and David Weenink (2004). *Praat: doing phonetics by computer.* http://www.fon.hum.uva.nl/praat/download_win.html.

Busby Peter A. and G. L. Plant (1995). "Formant frequency values of vowels produced by preadolescent boys and girls". In: *The Journal of the Acoustical Society of America* 97.4, pp. 2603–2606.

Clark, Herbert H. and Jean E. Fox Tree (2002). "Using uh and um in spontaneous speaking". In: *Cognition* 84, pp. 73–111.

DeJoy, Daniel A. and Gregory, Hugo H. (1985). "The relationship between age and frequency of disfluency in preschool children". In: *Journal of Fluency Disorders* 10, pp. 107–122.

Duez, Danielle (2001). "Acoustico-phonetic characteristics of filled pauses in spontaneous French speech: Preliminary results". In: *Proceedings of ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*, pp. 41–44.

Eklund, Robert (2004). *Disfluency in Swedish human–human and human–machine travel booking dialogues. Linköping Studies in Science and Technology.* Dissertation No. 882. Linköping: Unitryck, 388 p.

Esposito, Anna (2005). "Children's Organization of Discourse Structure Through Pausing Means. Nonlinear Analyses and Algorithms for Speech Processing". In: *Lecture Notes in Computer Science* 3817, pp. 108–115.

Farantouri, V., A. Potamianos, and S. Narayanan (2008). "Linguistic Analysis of Spontaneous Children Speech". In: *Proceedings of the 1st Workshop on Child, Computer and Interaction*. (sail.usc.edu/aigaion2/index.php/attachments/)

Fiestas, Christine E., Bedore, Lisa M., Peña, Elizabeth D. and Nagy, Vanessa J. (2005). "Use of mazes in the narrative language samples of bilingual and monolingual 4- to 7-year old children. " In Cohen, James, McAlister, Kara T., Rolstad, Kellie and MacSwan, Jeff (eds.): *Proceedings of the 4th International Symposium on Bilingualism*. Cascadilla Press, Somerville. pp. 730–740. (http://www.lingref.com/isb/4/057ISB4.PDF)

Furman, Reyhan and Aslı Özyürek (2007). "Development of interactional discourse markers: Insights from Turkish children's and adults' oral narratives". In: *Journal of Pragmatics* 39, pp. 1742–1757.

Gayraud, F., H. Lee, and M. Barkat-Defradas (2011). "Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects". In: *Clinical Linguistics and Phonetics* 25, pp. 198–209.

Gósy, Mária (2012). "BEA – A multifunctional Hungarian spoken language data base". In: *The Phonetician* 105−106, pp. 50−61.

Hooper, Celia R. and Ann Cralidis (2009). "Normal Changes in the Speech of Older Adults: You've still got what it takes; it just takes a little longer!" In: *Perspectives on Gerontology* 14, pp. 47–56.

Hudson Kam, Carla L. and Edwards, Nicole A. 2008. "The use of uh and um by 3- and 4-year-old native English-speaking children: Not quite right but not completely wrong." In *First Language* 28, pp. 313–327.

Ito, T. (1986). "Speech dysfluencies and acquisition of syntax in children 2–6 years old." In: *Folia Phoniatrica* 38, pp. 310.

Kemper, Susan. 1992. "Adults sentence fragments: Who, what, when, where and why? " *Communications Research* 19, pp. 444–458.

Lee, Sungbok, Potamianos, Alexandros and Narayanan, Shrikanth. (1999). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters. " In: *The Journal of the Acoustical Society of America* 105, pp. 1455–1468.

Leeper, Linda H., and Richard Culatta (1995). "Speech fluency: Effect of age, gender and context". In: *Folia Phoniatrica et Logopaedica* 47, pp. 1–14.

MacWhinney, Brian and Harry Osser (1977). "Verbal Planning Functions in Children's Speech". In: *Child Development* 48, 978−985.

Pickett, James M. (1980). *The sounds of speech communication*. Baltimore: University Park Press, 194 p.

Pindzola, Rebekah H. (1990). "Dysfluency characteristics of aged, normal-speaking black and white males". In: *Journal of Fluency Disorders* 15, pp. 235–243.

Roggia, Aaron B. (2012). "*Eh* as a polyfunctional discourse marker in Dominican Spanish. " In: *Journal of Pragmatics* 44, pp. 1783–1798.

Searl, J. P., R. M. Gabel, and J. S. Fulks (2002). "Speech disfluency in centenarians". In: *Journal of Communication Disorders* 35, pp. 383–392.

Shriberg, Elisabeth (1994). *Preliminaries to a Theory of Speech Disfluencies*. Ph.D. Dissertation, University of California.

Shriberg, Elizabeth (2001). "To „errrr" is human: ecology and acoustic of speech disfluencies". In: *Journal of the International Phonetic Association* 31, pp. 153–169.

Simpson, Adrian (2006). "Phonetic Processes in Discourse". In: Brown, Keith (ed.): *Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier, pp. 284–288.

Simpson, Adrian P. (2009). "Phonetic differences between male and female speech". In: *Language and Linguistics* 3, pp. 621–640.

Thordardottir, Elin T. and Weismer, Susan Ellis (2002). "Content mazes and filled pauses on narrative language samples of children with specific language impairment." In: *Brain and Cognition* 48, pp. 587–592.

Vorperian, Houri K. and Ray D. Kent (2007). "Vowel acoustic space development in children: a synthesis of acoustic and anatomic data". In: *Journal of Speech, Language, and Hearing Research* 50, pp. 1510–1545.

Yaruss, J. S., R. M. Newman, and T. Flora (1999). "Language and disfluency in non-stuttering children's conversational speech". In: *Journal of Fluency Disorders* 24, pp. 185–207.

Watanabe, M., K. Hirose, Y. Den, and N. Minematsu (2008). "Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners". In: *Speech Communication* 50, pp. 81–94.